Synthetic Range-Doppler Map Generating with WGAN-GP: A Study on Convergence and Image Realism for Radar Data Augmentation

Clara Reis Prata Baldansa, Ronaldo Ribeiro Goldschmidt, Julio Cesar Duarte

Instituto Militar de Engenharia (IME)
Praça Gen. Tibúrcio, 80 – 22290-270 – Urca, Rio de Janeiro - RJ – Brazil

{clara.baldansa,ronaldo.rgold,duarte}@ime.eb.br

Abstract. *Generating data for radar applications, such as range-Doppler maps, is often expensive and time-consuming, and available datasets are typically small, which limits the effectiveness of deep learning models and can significantly hinder performance in tasks like target classification. In this study, we analyzed the feasibility of using Generative Adversarial Networks (GANs), particularly a WGAN-GP with spectral normalization, to generate synthetic range-Doppler map images and expand data availability. Despite our efforts, the results showed that the WGAN-GP struggled to capture the dataset's complexity, with the discriminator being excessively stronger than the generator. This imbalance led to low visual quality in the generated images and poor performance in the Geometric Score (GS) and SSIM metrics. These findings highlight the challenges of training GANs in complex scenarios and the need for more robust methodologies for future data augmentation applications in radar systems.*

1. Introduction

In today's increasingly complex electromagnetic environments, with increased interference and noise [Wang et al. 2025], radar systems face growing challenges in target detection and classification [Slesinski and Wierzbicki 2025]. To overcome these difficulties, next-generation radar systems have begun incorporating deep learning technologies [Singh et al. 2025].

Convolutional Neural Networks (CNNs) have been successfully applied to target recognition tasks [Upadhyay et al. 2021], with their performance being influenced by the volume of training data [Luo et al. 2018]. Lack of training data, nevertheless, can lead to overfitting and, consequently, poor performance, and, in radar applications, the amount of labeled data available is often insufficient [Gusland et al. 2021].

According to [Jiang et al. 2022], who reviewed the use of deep learning for target detection, most research in this area relies on simulated data. This is largely due to the high operational cost and difficulty of recording multiple radar measurements in controlled environments. At the same time, there is a lack of computational modeling approaches that can consistently generate realistic results, and existing methods often exhibit high computational complexity [Holder et al. 2018].

In this context, this work investigates the feasibility of generating synthetic range-Doppler maps using generative adversarial networks, employing the Wasserstein distance as a loss function with gradient penalty regularization (WGAN-GP). The focus is on stabilizing the training process and analyzing the quality of the generated images in terms of realism, diversity, and their potential usefulness for augmenting radar datasets.

The results of this study revealed that WGAN-GP, although theoretically designed to mitigate training instability, failed to generate realistic range-Doppler map images. Evaluation metrics, such as Geometric Score GS (0.9665) and Structural Similarity Index SSIM (0.2297), pointed to the low quality and diversity of the synthetic images. This reflected a notable imbalance during training, in which the discriminator proved to be disproportionately strong, preventing the model from converging toward a satisfactory distribution.

2. Related Work

The study by [Park et al. 2022] employs a Conditional GAN (CGAN) to generate class-specific range-Doppler samples. The quality of the generated data is evaluated indirectly by training a CNN classifier with the augmented data, showing improved classification performance.

[Queiroz et al. 2021] uses a DCGAN and two evaluation metrics: one qualitative, based on visual inspection, and another quantitative, using the GS, which reached a value of 0.103. The author argues that, despite GS's limitations in capturing subtle visual aspects, the results indicate that GANs can generate potentially realistic radar samples.

The study by [de Oliveira and Bekooij 2020] applies a vanilla GAN; however, due to the inherently noisy nature of range-Doppler maps, this type of network is prone to convergence difficulties. To mitigate this problem, the author proposes a preprocessing step using Convolutional Autoencoders (CAE). To assess the quality of synthetic images, the generated data is used to train a detection system based on artificial neural networks. The system's performance is then tested using real images, providing a measure of the practical usefulness of the synthetic samples.

Other works [Lee et al. 2020, de Andrade et al. 2023] perform comparative analyses of different architectures, including CGAN, DCGAN, and WGAN-GP. [de Andrade et al. 2023] achieves the best performance for a GAN according to the "score combined" metric, while [Lee et al. 2020] shows that WGAN-GP outperformed CGAN based on the SSIM metric, with values of 0.6086, 0.4611, and 0.3308 for three different classes used in the experiment. These studies indicate that the optimal GAN architecture depends on multiple factors, including the underlying image distribution, chosen hyperparameters, and the evaluation metrics employed.

Unlike prior works, this study focuses on the challenges of generating range-Doppler maps using WGAN-GP with both gradient penalty and spectral normalization to enforce the Lipschitz constraint. Additionally, it provides a critical analysis of training instability, particularly examining how the imbalance between the generator and discriminator impacts the visual quality of generated images. Finally, the study evaluates image quality and diversity using GS and SSIM, which are more suited for radar data than metrics such as Fréchet Inception Distance (FID) [Heusel et al. 2017] and Inception Score (IS) [Salimans et al. 2016].

3. Methodology

This section explores the application of a WGAN-GP architecture with spectral normalization for generating synthetic range-Doppler maps. The main objective is to evaluate the feasibility of using these images for data augmentation in DL radar applications.

The Real Doppler RAD-DAR (RDRD) dataset [Roldan et al. 2020] was utilized in this study. It contains 17,493 labeled samples divided into three classes: *drone*, *car*, and *person*. The original images have dimensions of $(11 \times 61 \times 1)$ and were normalized to the range $[-1, 1]$ to align with the *tanh* activation function employed in the generator's output layer. To align the image dimensions with convolutional operations and optimize GPU execution, *zero padding* was applied, resulting in input images of size $(16 \times 64 \times 1)$. The generator outputs larger images and utilizes a final *Cropping2D* layer to match the original input size.

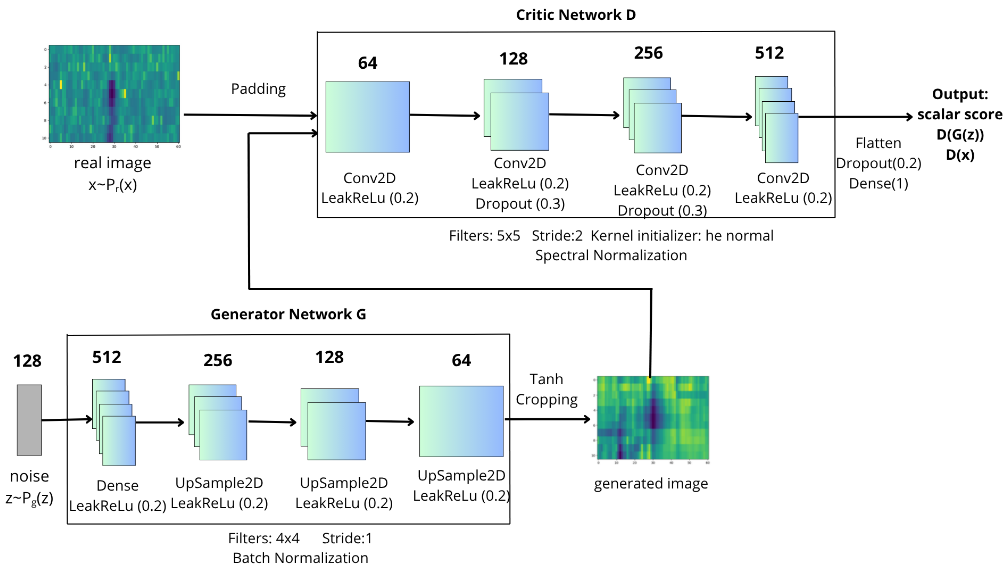


Figure 1. Architecture used in the experiment

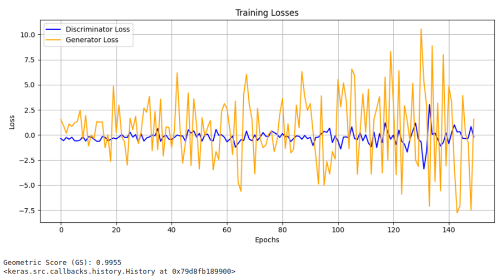
The latent space dimension was set to 128, sampled from a standard Gaussian distribution, and used as input to the generator. The generator network comprises a dense layer followed by four upsampling and convolutional blocks, with decreasing numbers of filters (512, 256, 128, 64, 1). All intermediate layers apply *Batch Normalization* and *LeakyReLU* activation with $\alpha = 0.2$, while the final layer uses *tanh*. The discriminator (or *critic*) consists of four convolutional layers with increasing numbers of filters (64, 128, 256, 512), all employing *LeakyReLU*, *Spectral Normalization*, and selective *dropout* with a 0.3 rate in intermediate layers. The output is a real-valued scalar without *sigmoid* activation, as prescribed in the original WGAN formulation. Figure 1 illustrates the complete architecture used in the experiment, including the number of layers, regularization techniques, and activation functions.

The experiments employed the Adam optimizer, with hyperparameters including the learning rate α and the coefficients for the first- and second-order moment estimates. The choice of hyperparameters was based on the work of [Gulrajani et al. 2017], which uses the following values for the hyperparameters: $\alpha = 1 \times 10^{-4}$, $\beta_1 = 0.5$, $\beta_2 = 0.9$,

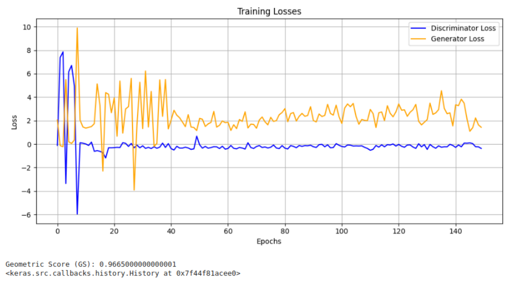
$\lambda = 10$, and $n_{\text{critic}} = 5$ (the number of training runs that the discriminator performs compared to the generator at each iteration). Additionally, the study by [Miyato et al. 2018] concludes that spectral normalization is relatively robust to variations in these parameters, so it was decided to maintain the initial settings of $\beta_1 = 0.5$, $\beta_2 = 0.9$, and $\lambda = 10$, concentrating efforts on other model improvements.

In preliminary tests, the discriminator quickly outperformed the generator. To mitigate this, the number of discriminator updates per training step was reduced from 5 to 3 ($n_{\text{critic}} = 3$), balancing the adversarial dynamics. The model was trained for 150 epochs, a value chosen empirically after observing that 50 epochs were insufficient for convergence, while 200 epochs resulted in discriminator overfitting. The batch size was set to 512. Spectral normalization was applied to all convolutional layers in the discriminator to enforce the 1-Lipschitz constraint in a stable manner. Although this technique was not sufficient to fully balance training, it improved training smoothness and the informativeness of gradients provided to the generator.

As shown in Figure 2a, in the absence of spectral normalization, the discriminator loss (blue curve) quickly converges and fluctuates around zero, while the generator loss (orange curve) remains unstable, indicating poor gradient feedback. In contrast, Figure 2b demonstrates that with spectral normalization, the discriminator loss decays more gradually, and the generator loss shows reduced oscillations. This adjustment led to a slight improvement in the GS (from 0.9955 to 0.9665), supporting the hypothesis that spectral normalization enhances stability and diversity, even if modestly.



(a) Discriminator without Spectral Normalization.



(b) Discriminator with Spectral Normalization.

Figure 2. Comparison of training behavior with and without spectral normalization.

Batch normalization was used in the generator to mitigate the effects of *covariate shift* during backpropagation, stabilize gradient flow, and reduce sensitivity to weight initialization [Ioffe and Szegedy 2015]. Conversely, batch normalization was avoided in the discriminator because the gradient penalty is computed on a per-sample basis, and introducing dependencies between samples via batch normalization can lead to inconsistencies in gradient estimation [Gulrajani et al. 2017].

Proper weight initialization is important for the convergence of DL models. Typically, *He initialization* is used for nonlinear activation functions, which is the case with LeakyReLu [He et al. 2015]. This strategy helps mitigate vanishing or exploding gradients, which are common in deep architectures. In practice, it contributed to training stability and improved convergence. In the generator, however, changing the initialization scheme (e.g., from *Glorot uniform* to *He normal*) did not result in significant improvements in training balance or stability. This outcome is likely due to the stabilizing effect of batch normalization, which reduces sensitivity to weight initialization.

To assess the quality of the generated range-Doppler images, we employed both qualitative inspection and two quantitative metrics: GS and SSIM. SSIM ranges from 0 (no similarity) to 1 (perfect match), providing a visual fidelity estimate relative to real samples. The SSIM was computed following the standard formulation proposed by [Wang et al. 2004]. On the other hand, GS evaluates the similarity between the topological structures of the real and generated data distributions, reflecting both diversity and coverage of the underlying data manifold [Khrulkov and Oseledets 2018]. Lower GS values indicate closer alignment to the true data distribution and better coverage of its variation. As these metrics assess complementary properties, both were adopted. Furthermore, unlike commonly used metrics such as IS and FID, which depend on large pretrained classifiers, SSIM and GS operate directly on the data. This makes them better suited for domain-specific datasets such as radar, where pretrained classifiers may not capture meaningful representations.

4. Results and Analysis

The evaluation metric values, GS and SSIM, are presented in Table 1. Comparatively, the GS calculated in the experiment was 0.9665, considerably higher than the reference value of 0.103 (as reported by [Queiroz et al. 2021]). In contrast, the obtained SSIM value of 0.2297 was relatively closer to the reference value of 0.3308 reported by [Lee et al. 2020] for their class 2, which contains more complex images due to variations in pixel intensity, noise, and a larger number of visual elements. However, direct comparison is limited due to differences in datasets and experimental settings.

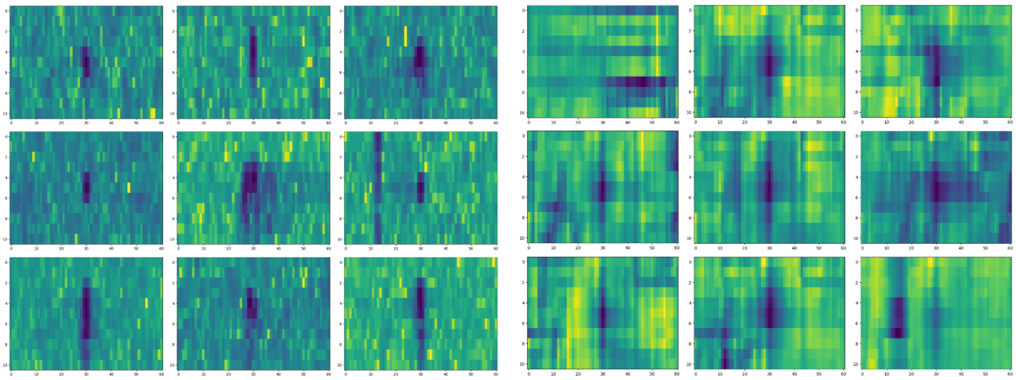
Metric	GS	SSIM
Reference	0.103	[0.6086, 0.4611, 0.3308]
Experiment	0.9665	0.2297

Table 1. Comparison of GS and SSIM in different experiments

In absolute terms, these results indicate unsatisfactory performance: a high GS suggests that the geometric distributions of real and generated images are significantly different, while a low SSIM reflects poor structural similarity. In short, both evaluation measures point to low quality and diversity, indicating that the model requires further adjustment.

From a qualitative perspective, based on visual inspection of real and generated images, it was observed that although some darker pixel regions representing similar target patterns could be identified, the generated images exhibited smoothed or “blurred” contours, lacking sharp structural details, as illustrated in Figure 3.

Despite the architectural choices described in Section 3, including attempts to mitigate gradient saturation using spectral normalization, gradient penalty, and weight initialization, the experimental results indicate that the WGAN-GP model failed to reach a stable adversarial equilibrium. Specifically, the persistent imbalance between the critic and generator losses throughout training, together with the poor quantitative and qualitative results, suggests that the critic progressively dominated the optimization process, limiting the generator’s ability to accurately approximate the target data distribution.



a) Real Images b) Generated Images

Figure 3. *Comparison between (a) real range-Doppler maps and (b) synthetic images generated by the WGAN-GP model.*

In summary, even by using WGAN-GP, which theoretically mitigates training instability, the model failed to reach equilibrium. This reinforces the need for further tuning, as GAN training is highly sensitive to the balance between networks [Heusel et al. 2017].

5. Conclusion and Future Work

DL has shown significant potential in enhancing radar signal processing, particularly in tasks involving automatic target detection and classification. These models rely heavily on large volumes of labeled data to learn discriminative features and generalize effectively. However, in radar domains, acquiring such data is often challenging due to the high cost and complexity associated with simulations and real-world data collection. Additionally, the scarcity of available datasets limits the development of robust models and increases the risk of overfitting.

To mitigate data limitations in radar applications, this study proposed using WGAN-GP to generate synthetic images from range-Doppler maps. The developed architecture incorporated gradient penalization, spectral normalization, and appropriate weight initialization in the discriminator, aiming to improve training stability and balance the adversarial process. While challenges remain in fully capturing the complexity of real radar signatures, the results provide valuable insights into the limitations of current generative models in this domain. The use of domain-specific evaluation metrics such as GS and SSIM also highlights the importance of fine-grained quality assessment, guiding future improvements in model design.

These results reinforce the challenges inherent in training GANs, such as sensitivity to hyperparameter selection and imbalance between the generator and the discriminator. Despite the regularization techniques adopted, the model did not converge to a satisfactory distribution, highlighting areas for architectural and configuration refinement. Nevertheless, these findings reflect important insights into the limitations of current GAN strategies in radar domains and contribute to a deeper understanding of how model stability and performance are affected by hyperparameter interactions in complex settings.

Future work may focus on automating hyperparameter selection through techniques such as Bayesian Optimization or Sequential Model-Based Optimization (SMBO). Additionally, complementary evaluation using metrics such as FID and IS. Another promising direction involves developing a conditional GAN capable of generating labeled range-Doppler maps, enabling targeted data augmentation and facilitating an analysis of its impact on the performance of supervised radar classification models.

6. Acknowledgments

The authors used a generative AI tool to improve the grammar and writing style. The AI tool was not used to generate scientific content.

References

de Andrade, C., Nogueira, M., Fidelis, M., Campos, L., Campos, P., Schön, P., and de Abreu Faria, T. (2023). Exploiting gan capacity to generate synthetic automotive radar data. In *Proceedings of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2023)*, volume 4: VISAPP, pages 262–271, Lisbon, Portugal. SciTePress.

de Oliveira, M. L. L. and Bekooij, M. J. G. (2020). Generating synthetic short-range fmcw range-doppler maps using generative adversarial networks and deep convolutional autoencoders. In *2020 IEEE Radar Conference (RadarConf20)*, pages 1–6, Florence, Italy. IEEE.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. (2017). Improved training of wasserstein gans. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 5769–5779, Red Hook, NY, USA. Curran Associates Inc.

Gusland, D., Christiansen, J. M., Torvik, B., Fioranelli, F., Gurbuz, S. Z., and Ritchie, M. (2021). Open radar initiative: Large scale dataset for benchmarking of micro-doppler recognition algorithms. In *2021 IEEE Radar Conference (RadarConf21)*, pages 1–6, Atlanta, GA, USA. IEE.

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1026–1034, Santiago, Chile. IEEE.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 6629–6640, Red Hook, NY, USA. Curran Associates Inc.

Holder, M., Rosenberger, P., Winner, H., D'hondt, T., Makkapati, V. P., Maier, M., Schreiber, H., Magosi, Z., Slavik, Z., Bringmann, O., and Rosenstiel, W. (2018). Measurements revealing challenges in radar sensor modeling for virtual validation of autonomous driving. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 2616–2622, Maui, HI, USA. IEEE.

Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR*, abs/1502.03167:448–456.

Jiang, W., Ren, Y., Liu, Y., and Leng, J. (2022). Artificial neural networks and deep learning techniques applied to radar target detection: A review. *Electronics*, 11(1).

Khrulkov, V. and Oseledets, I. (2018). Geometry score: A method for comparing generative adversarial networks. In *International Conference on Machine Learning*.

Lee, H., Kim, J., Kim, E. K., and Kim, S. (2020). Wasserstein generative adversarial networks based data augmentation for radar data analysis. *Applied Sciences*, 10(4).

Luo, C., Li, X., Wang, L., He, J., Li, D., and Zhou, J. (2018). How does the data set affect cnn-based image classification performance? In *2018 5th International Conference on Systems and Informatics (ICSAI)*, pages 361–366, Nanjing, China. IEEE.

Miyato, T., Kataoka, T., Koyama, M., and Yoshida, Y. (2018). Spectral normalization for generative adversarial networks. In *International Conference on Learning Representations*, Vancouver,Canada. OpenReview.net.

Park, S., Lee, S., and Kwak, N. (2022). Range-doppler map augmentation by generative adversarial network for deep uav classification. In *2022 IEEE Radar Conference (RadarConf22)*, pages 1–7, New York City, NY, USA. IEEE.

Queiroz, S. S. A., Shiguemori, E. H., Augusto, C. V. S., Souza, R. P., Damore, R., and Faria, L. A. (2021). Experimental assessment of the viability of a gan model for radar data generation. In *Simpósio de Aplicações Operacionais em Áreas de Defesa (SIGE)*, São José dos Campos - SP. SIGE.

Roldan, I., Del-Blanco, C. R., Duque de Quevedo, Á., Ibañez Urzaiz, F., Gismero Menoyo, J., Asensio López, A., Berjón, D., Jaureguizar, F., and García, N. (2020). Dopplernet: a convolutional neural network for recognising targets in real scenarios using a persistent range–doppler radar. *IET Radar, Sonar & Navigation*, 14(4):593–600.

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., and Chen, X. (2016). Improved techniques for training gans. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, page 2234–2242, Red Hook, NY, USA. Curran Associates Inc.

Singh, P., Menkudale, S., Soni, S., Tank, V. R., and Raj, A. A. B. (2025). Comprehensive survey on radar systems and its target classification techniques. *Journal of Electronic and Electrical Engineering*, 4(1):234–269.

Slesinski, J. and Wierzbicki, D. (2025). Review of synthetic aperture radar automatic target recognition: A dual perspective on classical and deep learning techniques. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 18:18978–19024.

Upadhyay, M., Murthy, S. K., and Raj, A. A. B. (2021). Intelligent system for real time detection and classification of aerial targets using cnn. In *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pages 1676–1681, Madurai, India. IEEE.

Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.

Wang, Z., Guo, Z., Shu, G., and Li, N. (2025). Radar jamming recognition: Models, methods, and prospects. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 18:3315–3343.